

Information spatiale de Lehmer et activation paramétrique pour la classification frugale de données positives

Matthieu GALLET¹ Abdourrahmane ATTO² Emmanuel TROUVÉ²

¹Centre de Recherches sur les Écosystèmes d'Altitude (CREA), Chamonix, FRANCE

²LISTIC (EA3703), Université Savoie Mont-Blanc, FRANCE

Résumé – Cet article introduit un formalisme d'apprentissage automatique fondé sur des cascades d'opérateurs paramétriques de Lehmer, généralisant les moyennes arithmétique et harmonique. Il repose sur un Filtre Spatial dit de Lehmer (FSL), qui extrait de manière non linéaire et apprenable des caractéristiques à partir de signaux. Pour classifier ces caractéristiques, une fonction d'activation adaptée est proposée. Testée sur des données suivant une loi de Rayleigh en imagerie SAR, cette approche améliore la performance en présence de peu d'exemples d'entraînement ou de labels bruités, notamment lorsque les informations classifiantes ne suivent pas une distribution gaussienne.

Abstract – This article introduces a machine learning framework based on cascades of parametric Lehmer operators, generalizing the arithmetic and harmonic means. It relies on a so-called Lehmer Spatial Filter (LSF), which extracts features from signals in a nonlinear and learnable manner. To classify these features, a suitable activation function is proposed. Tested on data following a Rayleigh distribution in SAR imaging, this approach improves performance when training examples are limited or labels are noisy, especially when classifying information that does not follow a Gaussian distribution.

1 Introduction

La plupart des méthodes de traitement du signal adoptent une approche basée sur les moyennes arithmétiques pour extraire des tendances ou d'autres caractéristiques supposées informatives en moyenne. Pourtant, des approches impliquant des moyennes géométriques [8] et des opérateurs logarithmiques [3] (hypothèses d'interactions multiplicatives affectant la tendance) ont été proposées afin de mieux se conformer à la nature de l'acquisition ou de l'observation de ces signaux, notamment en imagerie SAR [4], où le signal est perturbé par un bruit multiplicatif (chatoiement). Bien que les réseaux d'apprentissage profond compensent ces interactions par de très nombreuses cascades de convolutions et d'activations, certaines architectures exploitent des combinaisons multiplicatives [5], soulignant l'intérêt d'une approche pouvant être plus directe sur la représentation de certaines informations.

Le choix entre une approche arithmétique, harmonique ou géométrique (au sens *multiplicative*) reste ouvert et dépend sans doute surtout de la compacité qu'on désire assigner à l'opérateur extrayant les primitives utiles. L'opérateur de Lehmer [1] apporte une solution de type moyenne généralisée à ce problème d'assignation en couvrant un continuum entre les moyennes arithmétique et harmonique, entre autres. Il a montré son intérêt notamment en séparation de sources [7]. Son adaptation en activation convolutive de modèles d'apprentissage profond [6] répond au problème de l'agrégation sélective de caractéristiques, rejoignant des approches par apprentissage du *pooling* (rejet de valeurs ou sous-échantillonnage) optimal [2].

Nous proposons dans cet article une approche frugale où le modèle d'apprentissage n'extrait pas de nombreuses caractéristiques négatives (sorties convolutives quelconques par exemple) qu'il rejette ultérieurement à la phase d'extraction.

Cette approche repose uniquement sur des cascades d'opérateurs de Lehmer pour un apprentissage intrinsèquement adapté aux données strictement positives et à caractéristiques strictement positives. Nos contributions incluent la décomposition d'informations via des couches de Filtres Spatiaux de Lehmer (FSL) et la proposition d'une fonction d'activation paramétrique spécifique aux caractéristiques strictement positives. Après avoir défini le cadre théorique du FSL (sections 2 et 3), nous le comparons à la régression logistique et aux perceptrons multicouches à travers deux scénarios : impact de la quantité de données d'apprentissage et robustesse aux labels bruités (sections 4.1–4.2). Enfin, nous discuterons des performances et du potentiel du formalisme proposé (section 5).

2 Moyenne de Lehmer et Filtre Spatial de Lehmer (FSL)

Dans cette section, nous rappelons les notions existantes de moyenne de Lehmer et introduisons la formulation du Filtre Spatial de Lehmer bidimensionnel. Soit $\mathbf{x} \in (\mathbb{R}_+^*)^M$ un signal strictement positif de M échantillons. La moyenne de Lehmer est définie par :

$$L_p(\mathbf{x}) = \frac{\sum_{k=1}^M x_k^p}{\sum_{k=1}^M x_k^{p-1}} \quad (1)$$

Cette moyenne, de paramètre $p \in \mathbb{R}$, permet de balayer entre les opérateurs minimum et maximum de \mathbf{x} , respectivement pour $p = -\infty$ et $p = +\infty$. Elle inclut plusieurs cas particuliers, notamment la moyenne arithmétique pour $p = 1$ et la moyenne harmonique pour $p = 0$. De plus, il est important de noter que $L_p(\mathbf{x})$ est monotone (i.e. $p_1 \leq p_2 \leftrightarrow L_{p_1}(\mathbf{x}) \leq L_{p_2}(\mathbf{x})$) et que $L_p(\mathbf{x})$ reste dans l'intervalle $]0, +\infty[$. À par-

Ce travail a été partiellement financé par le Centre national d'études spatiales, France (CNES).

¹Les principaux opérateurs issus du FSL sont similaires aux opérateurs de pooling (moyenne, max, min) utilisés en apprentissage profond pour la réduction de dimension, tout en étant paramétriques.

tir de cette formulation, nous introduisons le concept de Filtre Spatial de Lehmer (FSL) d'ordre p , appliqué à une image carrée $\mathbf{X} \in (\mathbb{R}_+^*)^{M^2}$ avec un noyau carré de taille w , noté $\mathcal{H}_{p,w}$:

$$\mathcal{H}_{p,w}(\mathbf{X}) = \left\{ \frac{\sum_i^{i+w-1} \sum_j^{j+w-1} (X_{i,j})^p}{\sum_i^{i+w-1} \sum_j^{j+w-1} (X_{i,j})^{p-1}}, \quad (2) \right. \\ \left. i, j \in \{1, \dots, M - w + 1\} \right\}$$

Le noyau est uniquement utilisé pour définir la taille de la fenêtre d'analyse. La sortie de cet opérateur est une matrice $\mathcal{H}_{p,w}(\mathbf{X}) \in (\mathbb{R}_+^*)^{(M-w+1)^2}$. Nous proposons la structure de décomposition suivante. Étant donnée une image \mathbf{X} de taille (M, M) , nous appliquons une décomposition récursive utilisant le FSL avec une approche par fenêtre glissante et non-recouvrante sur un noyau de taille (w, w) . La décomposition de \mathbf{X} au niveau ℓ produit une carte de caractéristiques FSL de taille $(M^{\ell-1} - w + 1)^2$ au carrée. En pratique, la décomposition se poursuit jusqu'à L tel que $M = 1$, aboutissant à une seule caractéristique en entrée à ce niveau maximal pour représenter \mathbf{X} . Pour alléger la notation, nous n'explicitons plus le paramètre w , qui restera constant durant toute la décomposition. Le symbole \circ désigne la composition de deux opérateurs. Nous définissons le niveau L de décomposition de \mathbf{X} associé à la séquence de paramètres de Lehmer $\mathbf{P}_L = \{p_1, p_2, \dots, p_L\}$ par :

$$\mathbb{H}^{[\mathbf{P}_L]}(\mathbf{X}) = \mathcal{H}_{p_L} \circ \mathcal{H}_{p_{L-1}} \circ \mathcal{H}_{p_{L-2}} \circ \dots \circ \mathcal{H}_{p_1}(\mathbf{X}) \quad (3)$$

En pratique, nous choisirons la longueur L de la séquence de paramètres \mathbf{P}_L de sortes que $\mathbb{H}^{[\mathbf{P}_L]}(\mathbf{X}) \in \mathbb{R}_+^*$ soit un scalaire. Le niveau de décomposition correspondant est lié à la dimension de \mathbf{X} et à la taille du noyau considéré, et est donné par $L = (M - 1)/(w - 1)$.

L'apprentissage machine proposé consiste à apprendre \mathbf{P}_L , donc les L paramètres FSL associés à l'équation (3). Le modèle décrit par cette équation est explicable par le lien direct entre la valeur apprise de p_ℓ et la nature de la moyenne de Lehmer identifiée par apprentissage à ce niveau de décomposition $\ell \in \{1, 2, \dots, L\}$.

Dans la suite, nous aurons besoin de deux dérivées : celle de la ℓ -ième décomposition par rapport à la $\ell-1$ -ième, et celle de la ℓ -ième décomposition par rapport à son paramètre p_ℓ .

Pour faciliter la lecture, nous définissons :

$$S_q = \sum_{i=1}^{i+w-1} \sum_{j=1}^{j+w-1} \left(\mathbb{H}_{i,j}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X}) \right)^q,$$

$$T_q = \sum_{i=1}^{i+w-1} \sum_{j=1}^{j+w-1} \log \left(\mathbb{H}_{i,j}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X}) \right) \left(\mathbb{H}_{i,j}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X}) \right)^q,$$

avec (i, j) représentant les coordonnées de la fenêtre glissante du niveau ℓ et (h, m) les coordonnées de la fenêtre du niveau $\ell - 1$.

Les dérivées partielles s'écrivent alors :

$$\frac{\partial \mathbb{H}_{i,j}^{[\mathbf{P}_\ell]}(\mathbf{X})}{\partial p_\ell} = \frac{T_{p_\ell} S_{p_\ell-1} - S_{p_\ell} T_{p_\ell-1}}{S_{p_\ell-1}^2} \quad (4)$$

² $M^{\ell-1}$ étant la taille de l'image au niveau $\ell-1$ de la décomposition.

$$\frac{\partial \mathbb{H}_{i,j}^{[\mathbf{P}_\ell]}(\mathbf{X})}{\partial \mathbb{H}_{h,m}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X})} = \frac{p_\ell \left(\mathbb{H}_{h,m}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X}) \right)^{p_\ell-1} S_{p_\ell-1}}{S_{p_\ell-1}^2} \\ - \frac{(p_\ell - 1) \left(\mathbb{H}_{h,m}^{[\mathbf{P}_{\ell-1}]}(\mathbf{X}) \right)^{p_\ell-2} S_{p_\ell}}{S_{p_\ell-1}^2} \quad (5)$$

3 Activation positive paramétrique pour la classification

Afin d'effectuer une classification sur la décomposition proposée en Section 2 et ainsi obtenir une probabilité de classe, nous proposons d'ajouter une fonction d'activation. Les fonctions d'activation standards telles que la sigmoïde ou la tangente hyperbolique ne sont pas adaptées à notre cas, car elles sont conçues pour discriminer des caractéristiques positives *versus* négatives, alors que toutes les caractéristiques sont positives dans notre contexte. Nous proposons donc une fonction d'activation $\phi(x)$, qui doit satisfaire trois conditions :

- (1) être monotone sur \mathbb{R}_+^3
- (2) $\phi(x) \rightarrow 0$ lorsque $x \rightarrow 0$
- (3) $\phi(x) \rightarrow 1$ lorsque $x \rightarrow +\infty$

La méthode proposée introduit une agrégation récursive sans pondération, adaptée aux distributions positives. Cependant, la sortie de cette décomposition reste, par définition, dans l'intervalle $]\min(\mathbf{x}), \max(\mathbf{x})[$. Il est donc nécessaire de paramétrer la fonction d'activation afin de positionner le seuil de décision tel que $\phi(\delta) = 0.5$. Pour satisfaire les conditions (2) et (3), nous utilisons une fonction définie par morceaux avec un point d'inflexion en δ . Ce choix garantit que :

$$\left. \frac{\partial \phi(x)}{\partial x} \right|_{x \rightarrow \delta^-} = \left. \frac{\partial \phi(x)}{\partial x} \right|_{x \rightarrow \delta^+},$$

assurant ainsi la continuité de la première dérivée autour du point d'inflexion. Nous introduisons également un second paramètre k pour ajuster la pente de la fonction d'activation autour du seuil, permettant une meilleure séparation des classes dans le cas de distributions proches. Nous proposons ainsi une fonction d'activation paramétrique de la forme :

$$\phi(x) = \begin{cases} x^k & \text{si } x \leq \delta \\ \frac{2\delta^k}{1 + \exp(A)} & \text{sinon} \end{cases}, \quad (6) \\ \text{with } A = \frac{2k(\delta - x)}{\delta} \text{ et } \delta, k \in \mathbb{R}^+,$$

La construction détaillée de cette fonction n'est pas donnée ici par manque d'espace. La Figure 1 illustre la forme de cette fonction pour différentes valeurs de δ et k . En raison de la contrainte de positivité sur δ et k , nous proposons de les transformer sur \mathbb{R}_+ en définissant $\delta_+ = \exp(\delta)$ et $k_+ = \exp(k)$.

³ Nous assouplissons la contrainte de stricte positivité pour permettre une meilleure généralisation.

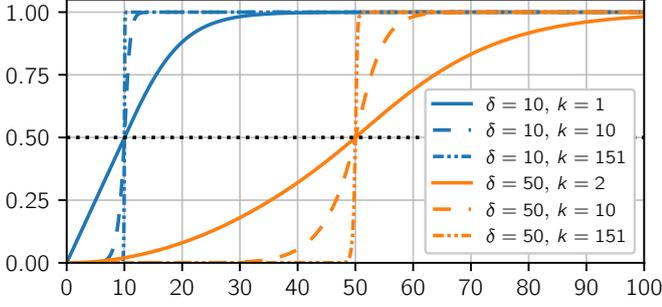


FIGURE 1 : Fonction d'activation paramétrique $\phi(x)$ pour différentes valeurs de δ et k

Les dérivées de $\phi(x)$ par rapport à x , δ et k sont données dans le Tableau 1, avec les coefficients B, C :

$$B = \frac{k_+}{2} \cdot x^{k_+-1} \cdot \delta_+^{-k_+} \quad (7)$$

$$C = \frac{\exp\left(\frac{2k_+}{\delta_+}(\delta_+ + x) - d + k\right)}{\left[\exp\left(2x \frac{k_+}{\delta_+}\right) + \exp(2k_+)\right]^2} \quad (8)$$

Soit la probabilité de classe estimée $\hat{y}_i = \phi \circ \mathbb{H}^{[P_L]}(\mathbf{X})$ et la vérité terrain associée y_i , la tâche de classification est réalisée avec une fonction de perte \mathcal{L} par entropie croisée sur N échantillons :

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N [-y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)] \quad (9)$$

Nous avons ainsi $L + 2$ paramètres à apprendre. En appliquant la règle de dérivation en chaîne, nous obtenons les gradients suivants pour δ_+ et k_+ :

$$\nabla_{\delta} = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \underbrace{\frac{\partial \hat{y}_i}{\partial \delta_+} \cdot \frac{\partial \delta_+}{\partial \delta}}_{\partial \phi / \partial \delta}, \quad \nabla_k = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \underbrace{\frac{\partial \hat{y}_i}{\partial k_+} \cdot \frac{\partial k_+}{\partial k}}_{\partial \phi / \partial k} \quad (10)$$

La dérivée analytique contrainte de $\partial \phi / \partial \delta$ et $\partial \phi / \partial k$ est donnée dans le Tableau 1. La dérivée standard $\partial \mathcal{L} / \partial \phi$ est donnée par :

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{1}{N} \frac{\hat{y}_i - y_i}{\hat{y}_i(1 - \hat{y}_i)} \quad (11)$$

La dérivée de \mathcal{L} par rapport à p_ℓ , le paramètre de la ℓ -ième décomposition, est obtenue de manière similaire. Avec l'opérateur (\cdot) représentant le produit matriciel.

$$\nabla_{p_\ell} = \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial \mathbb{H}^{[P_L]}(\mathbf{X})} \cdot \underbrace{\frac{\partial \mathbb{H}^{[P_L]}(\mathbf{X})}{\partial \mathbb{H}^{[P_L-1]}(\mathbf{X})}}_{\text{cf. Eq. 5}} \cdots \underbrace{\frac{\partial \mathbb{H}^{[P_\ell]}(\mathbf{X})}{\partial p_\ell}}_{\text{cf. Eq. 4}} \quad (12)$$

TABLE 1 : Dérivées de $\phi(x)$, par rapport à x , δ et k

Dérivée	$x \leq \delta_+$	$x > \delta_+$
$\frac{\partial \phi(x)}{\partial x}$	B	$2 \cdot C$
$\frac{\partial \phi(x)}{\partial \delta}$	$-Bx$	$-2 \cdot (\delta_+ - x) \cdot C$
$\frac{\partial \phi(x)}{\partial k}$	$Bx \cdot \log\left(\frac{x}{\delta_+}\right)$	$-2x \cdot C$

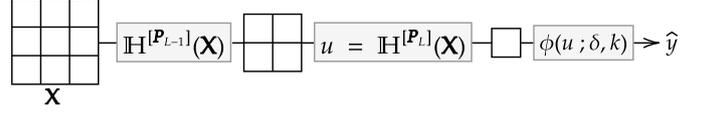


FIGURE 2 : Exemple de l'architecture proposée pour la classification des données positives, avec $w = 2$

Il est important de noter que la formulation $\partial \hat{y}_i / \partial \mathbb{H}^{[P_L]}(\mathbf{X})$ est donnée par la formule de dérivation $\partial \phi(x) / \partial x$ détaillée dans le Tableau 1. Enfin, les solutions pour les paramètres $\mathbf{P}_L = \{p_1, p_2, \dots, p_L\}$, δ et k sont obtenues via une descente de gradient classique. Une illustration complète de l'architecture proposée est donnée dans la figure 2.

4 Expériences numériques

Cette section illustre la capacité de l'architecture proposée à classifier des données simulées strictement positives. Nous considérons deux scénarios basés sur des distributions de Rayleigh \mathcal{R}^4 , définies sur \mathbb{R}^+ et couramment utilisées pour modéliser les images SAR en amplitude à une seule vue [9]. Leur paramétrisation unique simplifie la comparaison entre configurations. Ces tests simples visent à valider le concept. Dans les deux scénarios étudiés, nous générons deux classes \mathcal{C}_1 et \mathcal{C}_2 , paramétrées par σ_1 et σ_2 . L'ensemble de données comprend N échantillons de dimension 13×13 , répartis équitablement entre les classes, avec 80% pour l'entraînement et 20% pour le test. Chaque scénario est évalué sur 20 entraînements indépendants. Nous considérons deux tailles de noyaux : $w = 5$ et $w = 7$, induisant respectivement 3 et 2 niveaux de décomposition, soit un total de 5 et 4 paramètres à apprendre. Les performances sont comparées à celles d'un perceptron multicouche (MLP) et d'une régression logistique (LR) définie par :

$$\hat{y} = \sigma(\mathbf{w}^T \mathbf{x} + b),$$

où σ est la fonction sigmoïde⁵. L'entrée est vectorisée, nécessitant l'apprentissage de 170 paramètres pour la LR. Le MLP comporte deux couches cachées de taille 50 et 10, totalisant 9021 paramètres. Aucune contrainte n'est imposée sur la régularisation ou la positivité des poids. L'évaluation se fait à l'aide du F1-score pondéré.

4.1 Sensibilité au nombre de données d'entraînement

Nous évaluons les performances de classification pour $\sigma_1 = 1$ et $\sigma_2 = 1.1$, en faisant varier le nombre d'échantillons d'entraînement de 20 à 1500 (Figure 3). Jusqu'à 1000 échantillons, la solution proposée surpasse la régression logistique, avec un écart maximal de 30% à 250 échantillons. Les deux tailles de noyaux donnent des performances similaires, avec un léger avantage pour $w = 7$ (+3%). La régression logistique perd près de 30% de performance en dessous de 700 échantillons, mais devient plus efficace au-delà de 1000 (< 2% d'écart). Le MLP suit une tendance proche de la régression logistique sous 500 échantillons, puis progresse lentement. La solution proposée est plus stable et performante pour un faible nombre d'échantillons, grâce à un nombre réduit de paramètres à apprendre.

⁴ $\mathcal{R}(x; \sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$, $x \geq 0$, $\sigma > 0$.

⁵Avec \mathbf{w} le vecteur de poids et b le biais.

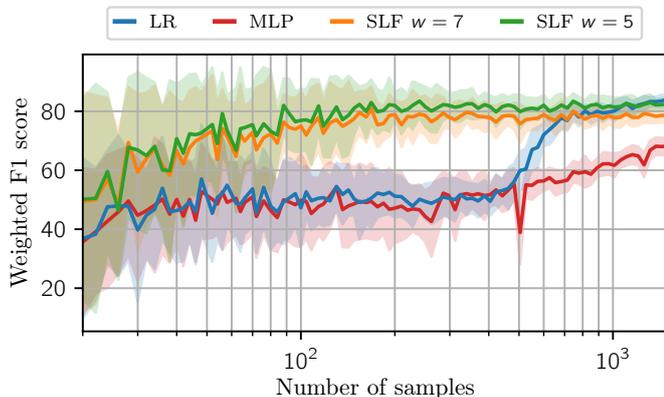


FIGURE 3 : F1-score pour la classification des échantillons de \mathcal{C}_1 et \mathcal{C}_2 pour différents nombres d'échantillons d'entraînement, avec $\sigma_1 = 1$ et $\sigma_2 = 1.1$.

4.2 Sensibilité au bruit de label

En conservant les mêmes paramètres σ_1, σ_2 que dans la section 4.1, nous analysons la sensibilité des deux méthodes au bruit sur les labels. Un pourcentage de labels a été inversé aléatoirement entre les deux classes, variant de 0% à 50%. Les résultats sont présentés dans la Figure 4.

La solution proposée surpasse la régression logistique pour tous les niveaux de bruit sur les labels, avec une variation moindre. Cependant, bien que les performances avec $w = 5$ et $w = 7$ soient respectivement supérieures de 16% et 19% à 5% de bruit, elles diminuent plus vite avec l'augmentation du bruit. À 50% de bruit, les performances chutent de 9% et 10% pour $w = 5$ et $w = 7$, contre 5% pour la régression logistique, atteignant 63% et 65% contre 56% pour la régression logistique. Le MLP reste le moins performant, avec des résultats entre 53% et 50%. Dans tous les scénarios, FSL avec $w = 7$ est plus performant que $w = 5$, car une fenêtre plus large permet une estimation locale plus précise. Toutefois, en contexte réel, des noyaux plus petits pourraient être complémentaires. Enfin, FSL surpasse la régression logistique, qui elle-même dépasse le MLP, bien que FSL ait moins de paramètres à apprendre. Cela souligne un effet de surapprentissage pour les deux autres méthodes.

5 Conclusion

Dans ce travail, nous avons conçu un modèle de réseau pyramidal basé sur des stratégies de filtrage non linéaire de Lehmer. Les variables apprenables incluent les paramètres des opérateurs de Lehmer et ceux d'une fonction d'activation adaptée. Cette approche, avec moins de paramètres que la régression logistique, conserve le domaine de définition des données tout au long de la décomposition et repose sur des opérateurs similaires au pooling en apprentissage profond. Pour la classification, nous avons proposé une nouvelle fonction d'activation paramétrée. Le protocole expérimental simple sur données simulées vise principalement à démontrer la faisabilité technique de l'implémentation et la réalisation concrète de notre méthode d'apprentissage automatique sans a priori de distribution, tout en préservant une partie de la structure spatiale des données de manière frugale et explicable. Les tests sur l'impact du nombre d'échantillons et la robustesse aux labels bruités montrent une nette amélioration, particulièrement pour les distributions po-

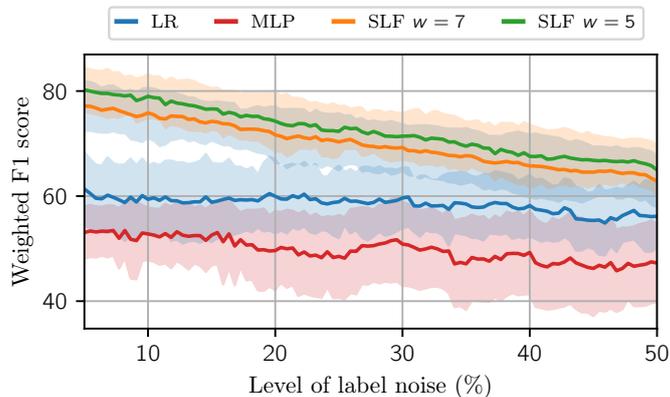


FIGURE 4 : F1-score pour la classification des échantillons de \mathcal{C}_1 et \mathcal{C}_2 en fonction des niveaux de bruit sur les labels, avec $\sigma_1 = 1$ et $\sigma_2 = 1.1$.

sitives et les jeux de données avec peu d'échantillons. Les perspectives prometteuses incluent l'extension vers des structures 2D et multidimensionnelles complexes où l'identification de distributions paramétriques appropriées devient difficile, positionnant notre approche comme alternative aux couches neuronales traditionnelles pour un apprentissage multi-échelle.

Références

- [1] Derrick H LEHMER : On the compounding of certain means. *Journal of Mathematical Analysis and Applications*, 36(1):183–200, 1971.
- [2] Brian MCFEE, Justin SALAMON et Juan Pablo BELLO : Adaptive pooling operators for weakly labeled sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(11):2180–2193, 2018.
- [3] Michael C. NEWMAN : Regression analysis of log-transformed data : Statistical bias and its correction. *Environmental Toxicology and Chemistry*, 12:1129–1133, 1993.
- [4] Guillaume QUIN, Béatrice PINEL-PUYSSÉGUR, Jean-Marie NICOLAS et Philippe LOREAUX : Mimosa : An automatic change detection method for sar time series. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9):5349–5363, 2014.
- [5] Mingxing TAN et Quoc V. LE : Efficientnet : Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.
- [6] Vagan TERZIYAN, Diana MALYK, Mariia GOLOVIANKO et Vladyslav BRANYTSKYI : Hyper-flexible convolutional neural networks based on generalized lehmer and power means. *Neural Networks*, 155:177–203, 2022.
- [7] Karn WATCHARASUPAT, Anh H. T. NGUYEN, Ching-Hui OOI et Andy W. H. KHONG : Directional sparse filtering using weighted lehmer mean for blind separation of unbalanced speech mixtures. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4485–4489, 2021.
- [8] Dennis WEI, Karthikeyan Natesan RAMAMURTHY, Dmitriy A. KATZ-ROGOZHNIKOV et Aleksandra MOJSILOVIĆ : Multiplicative regression via constrained least squares. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 304–307, 2014.
- [9] Dong-Xiao YUE, Feng XU, Alejandro C. FRERY et Ya-Qiu JIN : Synthetic aperture radar image statistical modeling : Part one—single-pixel statistical models. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):82–114, 2021.